

Quantum state tomography: Mean squared error matters, bias does not

Jiangwei Shang,¹ Hui Khoo Ng,^{1,2} and Berthold-Georg Englert^{1,3}

¹Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore

²Yale-NUS College, 6 College Avenue East, Singapore 138614, Singapore

³Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542, Singapore

(Dated: May 22, 2014)

Because of the constraint that the estimators be *bona fide* physical states, any quantum state tomography scheme—including the widely used maximum likelihood estimation—yields estimators that may have a bias, although they are consistent estimators. Schwemmer *et al.* (arXiv:1310.8465 [quant-ph]) illustrate this by observing a systematic underestimation of the fidelity and an overestimation of entanglement in estimators obtained from simulated data. Further, these authors argue that the simple method of linear inversion overcomes this (perceived) problem of bias, and there is the suggestion to abandon time-tested estimation procedures in favor of linear inversion. Here, we discuss the pros and cons of using biased and unbiased estimators for quantum state tomography. We conclude that the little occasional benefit from the unbiased linear-inversion estimation does not justify the high price of using unphysical estimators, which are typically the case in that scheme.

PACS numbers: 03.65.Ud, 03.65.Wj, 06.20.Dk

I. INTRODUCTION

The goal of quantum state tomography, or quantum state estimation, is to arrive at a best guess of the unknown state ρ of a quantum system, based on data collected from measuring a number of identical copies of the state. An accurate guess is needed in all aspects of quantum information or quantum computation, ranging from the characterization of an unknown quantum communication channel, to a check of a quantum gate implementation, or to the verification of a state preparation procedure in the lab.

Whether the guess from a particular tomography recipe can be considered the best, or most accurate, depends on one's figure-of-merit, which should be chosen according to the quantum information processing task at hand. In many situations, one is interested not in the state of the system itself, but in a quantity computed from it, e.g., the amount of entanglement in the state. In such cases, rather than reporting a best guess for ρ , one expects to get a more accurate answer by directly estimating the quantity of interest from the data, as done in a related procedure carrying the name of “parameter estimation.” In other situations, one is interested in a range of quantities related to ρ , and reporting a best guess for the state itself [e.g., by maximizing the likelihood for the data over all physical states, as is done for the maximum-likelihood estimator (MLE)] can be a convenient and self-consistent way of summarizing and interpreting the data.

In the latter case, one might assess the accuracy of the estimate obtained from a particular tomography scheme by examining some measure of closeness between the estimate $\hat{\rho}$, and the true state ρ , for a variety of known true states (e.g., from sources that have previously been fully characterized). A poorer, but possibly useful, gauge is to look at the accuracy of the prediction of one of the quantities of interest computed from $\hat{\rho}$.

Reference [1] compares the performance between an estimator $\hat{\rho}_{\text{LIN}}$ from a procedure the authors refer to as linear inversion (LIN), and two other estimators, $\hat{\rho}_{\text{MLE}}$ (the standard

MLE [2, 3]), and $\hat{\rho}_{\text{FLS}}$ (from a procedure known as “free least squares” (FLS) [4]). The article assesses the three estimators by looking at the accuracy in the prediction for target fidelity, a relevant quantity when the source is supposed to produce a certain target state.

The fidelity $F(\rho_1, \rho_2) \equiv \text{Tr}\{\sqrt{\rho_1}\sqrt{\rho_2}\}^2$ takes value between 0 and 1 for physical ρ_1 and ρ_2 [5]. In Ref. [1], the quality of an estimator is measured by the target fidelity $F(\rho_T, \hat{\rho})$, the fidelity between the target state ρ_T and the estimator $\hat{\rho}$, which is compared with the actual true value $F_0 \equiv F(\rho_T, \rho)$ computed for the true state ρ . Here, the “true” state yields the probabilities that are used for the generation of the simulated data from which the various estimators are derived.

The authors of Ref. [1] perform this comparison for different ρ_T s and ρ s, for many repeated simulations of the measurement data (and hence, many $\hat{\rho}$ s, one for each data). They draw the conclusion that $\hat{\rho}_{\text{LIN}}$ is always the best, because it is an unbiased estimator, i.e., fluctuations from different runs of the same experiment lead to fluctuations in the predicted F value, but all centered about the true value F_0 ; $\hat{\rho}_{\text{MLE}}$ and $\hat{\rho}_{\text{FLS}}$, on the other hand, give predictions that are biased, i.e., have a systematic shift away from F_0 (see Fig. 1 of Ref. [1] and Fig. 2 below). The authors go on to point out that any estimation procedure that always produces a physical (i.e., non-negative) state will unavoidably be biased; their $\hat{\rho}_{\text{LIN}}$, coming from an unbiased estimation procedure, is not guaranteed to be a physical density operator, and in fact, generically has negative eigenvalues [1].

Here, the qualifiers *biased* and *unbiased* have the technical meaning that is discussed below in the context of Eq. (2). Contrary to their connotation in common parlance, they are not synonyms of “bad” and “good.” One must not fall into the trap of regarding a biased estimator as automatically inferior to an unbiased one.

Indeed, it is well known in classical statistics that unbiased estimators are not always the best choice. Instead, minimizing the mean squared error (MSE, a popular measure of estimation accuracy) is key, and this is often not accomplished by

minimizing the bias. In fact, we will show that the LIN approach yields MSEs that are comparable to (and sometimes worse than) what one obtains from the MLE procedure; yet the LIN technique forces us to give up physicality, which leads to many severe problems and highly restricts the usefulness of the estimator $\hat{\rho}_{\text{LIN}}$. The MLE itself also does not—and was never designed to—minimize the MSE, but it does a comparably good job as $\hat{\rho}_{\text{LIN}}$ while enforcing physicality. We hence see little utility at all in employing the LIN strategy.

Below, we remind the reader why, in the quantum context, it is usually not a good idea to treat relative frequencies obtained from the data as probabilities, as is prescribed by the LIN procedure of Ref. [1]. Then, we explain why focusing on reducing bias only, and not the overall MSE, constitutes a conceptual misunderstanding. Lastly, we compare the MSEs obtained from the MLE and the LIN approaches and observe that one can easily find examples in which the biased MLEs have smaller MSEs than the unbiased LIN estimators.

II. FREQUENCIES ARE NOT PROBABILITIES

Before we begin, a brief note on notation is in order. The tomography measurement is described by a positive-operator-valued measure (POVM), or, if we use a more descriptive name, a probability-operator measurement (POM): It comprises a set of outcomes M_k , one for each detector, with $M_k \geq 0$ for all k and $\sum_k M_k = 1$. The probability of getting a click in detector k , corresponding to outcome M_k , is given by the Born rule, $p_k = \text{Tr}\{\rho M_k\}$. In the tomography experiment, N identically prepared copies of the (unknown) state ρ are measured using the POM. The data D consist of a particular sequence of detector clicks, summarized by the set of relative frequencies $f_k = n_k/N$, where n_k is the number of clicks in detector k . From D , one estimates the probabilities \hat{p}_k using a chosen procedure like MLE or LIN, and from these (if one can, e.g., in the case of tomographically complete POMs), one constructs the estimator $\hat{\rho}$.

LIN, as proposed in Ref. [1], sets the estimated probabilities equal to the relative frequencies of the observed data,

$$\hat{p}_k = f_k, \quad (1)$$

and then obtains $\hat{\rho}_{\text{LIN}}$ by “linear inversion” of the Born rule $\hat{p}_k = \text{Tr}\{\hat{\rho}_{\text{LIN}} M_k\}$. While relative frequencies will be close to probabilities when there is a lot of data, they are most certainly not the same thing: Relative frequencies satisfy only one constraint, that of unit sum: $\sum_k f_k = 1$; probabilities p_k (for POM $\{M_k\}$) that arise from a physical state ρ through the Born rule satisfy further constraints imposed by the positivity of ρ . The latter constraints can be easily stated for the case of measuring a qubit state with the symmetric informationally complete POM (SIC POM), the tetrahedron measurement [6, 7], where $\rho \geq 0$ requires the four tetrahedron probabilities to satisfy $\sum_k p_k^2 \leq \frac{1}{3}$, in addition to the unit-sum constraint. Since the relative frequencies do not themselves satisfy these physicality constraints, $\hat{\rho}_{\text{LIN}}$ is hence not necessarily a physical state, as is also emphasized in Ref. [1] (and many other existing references in the literature).

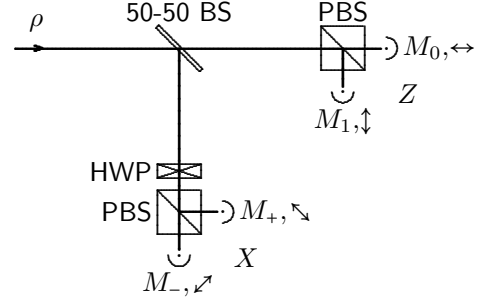


FIG. 1: X and Z measurements of a single qubit. The set-up uses a beam splitter (BS), a half-wave plate (HWP), two polarizing beam splitters (PBSs), and four detectors.

That $\hat{\rho}_{\text{LIN}}$ is not necessarily nonnegative, is not a minor nuisance: Many quantities associated with a physical state ρ are ill-defined for $\hat{\rho}_{\text{LIN}} \not\geq 0$ and can no longer be computed, e.g., entropy, negativity, and the fidelity with another state. Other quantities, such as the purity $\text{Tr}\{\hat{\rho}_{\text{LIN}}^2\}$ or the expectation value $\text{Tr}\{A\hat{\rho}_{\text{LIN}}\}$ of an observable A , are computable for $\hat{\rho}_{\text{LIN}} \not\geq 0$, but the numbers so obtained do not mean purity, expectation value, etc. Hence, $\hat{\rho}_{\text{LIN}}$ may not just lack a reasonable physical interpretation, but may also not be useful at all. In the case of uncomputable quantities, the proposal of Ref. [1] is to be content with the bounds that can be computed from linear approximations. These bounds, however, also lack a physical meaning if they are evaluated for $\hat{\rho}_{\text{LIN}} \not\geq 0$.

While one might choose not to be too concerned if $\hat{\rho}_{\text{LIN}}$ is only slightly unphysical (however one may want to quantify that statement), or if an unphysical $\hat{\rho}_{\text{LIN}}$ occurs only rarely, getting an unphysical $\hat{\rho}_{\text{LIN}}$ can be generic in certain situations. For example, imagine a qubit state measured with the tetrahedron measurement, and suppose that the true state is orthogonal to one of the tetrahedron outcomes (say the one labeled by $k = 1$). Then, the only relative frequencies that can give a physical $\hat{\rho}_{\text{LIN}}$ are $f_1 = 0$ and $f_2 = f_3 = f_4 = \frac{1}{3}$, i.e., the detector counts for all outcomes, other than the tetrahedron leg orthogonal to the true state, must be exactly equal. This is not even possible if the total number of counts is different from a multiple of 3.

Lest the reader complains that the above is a pathological case, another situation where one sees a stark contrast between frequencies and probabilities can be found in the commonly used “BB84-like” measurements, i.e., measure the Pauli X and Z on a single qubit with equal probability. In an optical implementation, where the qubit is the photon polarization, the usual way this measurement is implemented is by having a 50-50 beam-splitter direct the incoming photons into two possible paths, one carrying out the X measurement, the other the Z measurement; see Fig. 1. Now, the probabilities for such a measurement, by the very nature of the measurement structure, satisfy $p_0 + p_1 = p_+ + p_- = \frac{1}{2}$, alongside the positivity constraint $(p_0 - p_1)^2 + (p_+ - p_-)^2 \leq \frac{1}{4}$. The relative frequencies, however, obey no such constraints: Despite the

50-50 nature of the beam-splitter, one hardly ever encounters the situation where *exactly* half the photons travel down the X path, and half down the other. The procedure of finding $\hat{\rho}_{\text{LIN}}$ from such relative frequencies will then typically be internally inconsistent, and yields no solution.

One common fix used to circumvent the above problem requires one to ignore the counts in one of the detectors, e.g., the one measuring the $+1$ eigenstate for X (outcome M_+ in Fig. 1) [4]. To ensure that the relative frequencies comply with the constraint of $f_0 + f_1 = f_+ + f_- = \frac{1}{2}$, in imitation of the probabilities, one replaces the actual count n_+ obtained by $n_+ = n_0 + n_1 - n_-$, and modifies the total number N to be $2(n_0 + n_1)$. However, this ad-hockery, which involves discarding data, does not guarantee that the relative frequencies satisfy the remaining positivity constraint on the probabilities.

The simulations in Ref. [1] mimic the tomography of four-qubit states using product Pauli POMs, i.e., measure $[O_1 O_2 O_3 O_4] \equiv O_1 \otimes O_2 \otimes O_3 \otimes O_4$, with each of the four O_α s equal to one of the three Pauli operators X , Y , and Z . Rather than just having two settings of X and Z as in the single-qubit case of Fig. 1, there are now $3^4 = 81$ settings, all to be measured with equal probability. A practical way of carrying out this tomography measurement, having (usually) no access to a 1-to-81 beam-splitter, would be to divide the total number of counts by 81, and measure $N/81$ copies of ρ with each setting (this is another way of overcoming, by hand, the difficulty discussed in the previous two paragraphs). This automatically ensures that the relative frequencies for each setting $[O_1 O_2 O_3 O_4]$ sum to $1/81$, as do the corresponding probabilities.

However, these constraints are not the only ones needed to ensure internal consistency. There is, for example, the problem that frequencies obtained when measuring, say, $[XZZY]$ and $[XYXY]$ give two values for the expectation values of the two-qubit observable $[X_{\text{---}}Y] = X \otimes 1 \otimes 1 \otimes Y$ and also for the single-qubit observables $[X_{\text{---}}]$ and $[_{\text{---}}Y]$. In Ref. [1], these conflicts are resolved by a fix that may appear plausible, but is ad-hockery nevertheless and ultimately difficult to justify. Furthermore, there is still the issue of positivity constraints, and our own simulations show that almost all the estimators constructed by LIN violate positivity and, therefore, are unphysical.

None of these ad-hoc fixes or discard of data are necessary if we do not insist on setting frequencies equal to probabilities. The usual approach of finding a *physical* density operator that best fits the data, e.g., by maximizing the likelihood for the data over all physical states in the case of the MLE, automatically handles all these constraints. There is no need to include constraints explicitly by hand, such a need becoming more severe and difficult to carry out as the dimensionality of the system and the complexity of the POM increase. Of course, finding the best-fit physical estimator for the data is also not easy, but the estimator is assuredly physical, as is the true state, and one interprets the data honestly without additional tweaks.

III. TO BIAS OR NOT TO BIAS?

The accuracy of an estimator is often quantified by the mean-squared error (MSE), i.e., the mean, over all possible data, of the squared deviation of the estimated parameter from the true value, but this is only one of many measures of inaccuracy. To be concrete, we take as example the estimate \hat{F} for the target fidelity (with respect to some target state ρ_T) as discussed in Ref. [1]. The MSE in this case is

$$\mathbb{E}_{\rho_T}[(\hat{F} - F_0)^2] = [\mathbb{E}_{\rho_T}(\hat{F}) - F_0]^2 + \mathbb{V}_{\rho_T}(\hat{F}), \quad (2)$$

where we adopt the notation of Ref. [1]: \mathbb{E}_{ρ_T} denotes the expected value and \mathbb{V}_{ρ_T} the variance, for the true state ρ_T . The MSE in Eq. (2) is the sum of two contributions: The first is the so-called *bias* of the estimator (as compared with the true value F_0); the second is the variance of \hat{F} . As it is written, the formula for the MSE makes no statement about the relative importance of the two pieces, and a general estimation strategy can have different relative sizes for, and compromises between, the two terms. Other measures of inaccuracy may not have a break-up analogous to Eq. (2), and the bias may not be a relevant notion.

It is easy to understand intuitively why both pieces, not just the bias, must be small in order for us to have reasonable confidence of obtaining a good estimate from a single set of data. If the bias is small, and one has many runs of the same experiment, yielding very many estimates, one might imagine getting a good answer by looking for the centre of all these estimates. However, one usually has access to only one set of data, and hence a single estimate [8]; other estimates may then be generated by bootstrapping the data, but this just gives a distribution of values centered at the data-based estimate. If the estimation strategy yields no bias, but a large variance, one will often end up with an estimate that is quite far away from the true value. Hence, one needs both the bias, as well as the variance, summarized as the MSE, to be small for an accurate guess.

Most statistics textbooks (see, for example, [9]) do focus on the class of unbiased estimators for a given problem. However, this is always discussed together with the search for one with the minimum variance, i.e., to minimize the MSE *within the class of unbiased estimators*. The resulting theory of point estimation for unbiased estimators has rather nice and concise results, with wide applicability in many areas of statistics. Yet, at the same point where unbiased estimators are discussed, textbooks usually remind readers that unbiased estimators are not the whole story. The constraint of unbiasedness is often too strong: An estimator with a slight bias but a significantly smaller variance can very well be the estimator that minimizes the MSE. Certainly, considering only unbiasedness without also trying to minimize the variance, as suggested in Ref. [1], is altogether insufficient to give a good estimator.

Furthermore, as most textbooks will also explain, reasonable unbiased estimators do not exist in many problems. The nonexistence of an unbiased estimator that satisfies the positivity constraint for quantum states, as is emphasized in Ref. [1], is but another example of the difficulty in requiring

unbiasedness in constrained estimation problems. As a result of constraints, pure states and other rank-deficient states are extremal and the valid (*i.e.*, satisfy the constraints) estimators cannot approach them from all sides: The estimators are unavoidably biased. This is clearly visible in the examples discussed in the next section.

The trustworthy estimators, such as the MLEs, are *consistent*, *i.e.*, they converge to the true state when more and more data are taken into account. This consistency is an important property that is worth insisting upon, while unbiasedness is not.

The various virtues of, and problems with, unbiased estimators have been repeatedly expounded upon in the statistics literature. As an example, we point the reader to the excellent discussion found in Jaynes's book on probability theory, Ref. [10] (see Chapter 17). Within Jaynes's book is an example where the requirement of unbiased estimators leads one to waste half the data as compared with a biased estimator that attains the same MSE [11].

IV. MSE FOR LIN AND MLE APPROACHES

To illustrate the issues mentioned in the previous sections, we re-examine the example of target fidelity discussed in Ref. [1]. The unknown true state is a four-qubit state, and the parameter of interest is its fidelity with some target state ρ_T . The measurement comprises the 81 product Pauli operators described above, with 100 copies measured per setting, and 16 possible outcomes per setting. To explore the statistics, for each set of target and true states, we simulate 500 runs of the experiment, obtain the estimators $\hat{\rho}_{\text{MLE}}$ and $\hat{\rho}_{\text{LIN}}$ for each set of data, and from these, estimate the target fidelity.

As observed above, there is a problem with evaluating (not to mention, interpreting the meaning of) the target fidelity $F(\rho_T, \hat{\rho}_{\text{LIN}}) = \text{Tr}\{|\sqrt{\rho_T}\sqrt{\hat{\rho}_{\text{LIN}}}|^2\}$ when $\hat{\rho}_{\text{LIN}}$ is not non-negative. Reference [1] also admits this problem, and suggests to fix it by using the formula $F(\rho_T, \hat{\rho}_{\text{LIN}}) = \text{Tr}\{\rho_T \hat{\rho}_{\text{LIN}}\}$ instead, which is correct if $\hat{\rho}_{\text{LIN}}$ is nonnegative and ρ_T is a pure state. Although $\text{Tr}\{\rho_T \hat{\rho}_{\text{LIN}}\}$ can be computed also for an unphysical $\hat{\rho}_{\text{LIN}}$, it does not have the significance of the target fidelity then. The other issue, *viz.* that ρ_T must be pure, is of lesser concern if all target states of interest are indeed pure; the Smolin state considered in Ref. [1] is not pure, and this is why the respective line for $\hat{\rho}_{\text{LIN}}$ is missing in their Fig. 2. Clearly, one already sees the problems with using a nonphysical estimator.

Nevertheless, we can still compare the estimated target fidelities obtained from the MLE and LIN procedures for pure target states only, using their suggested formula of $F(\rho_T, \hat{\rho}) = \langle \psi | \hat{\rho} | \psi \rangle$ for $\rho_T = |\psi\rangle\langle\psi|$, whether $\hat{\rho}$ is physical or not. As the first target state, we consider the four-qubit Greenberger-Horne-Zeilinger (GHZ) state with the ket $|\text{GHZ}_4\rangle = \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$. This is the main example discussed in Ref. [1]; we simulate the identical situation and the results are shown in Fig. 2, which are very similar to those in Fig. 1 of Ref. [1].

Instead of mixing the GHZ state with white noise, we can

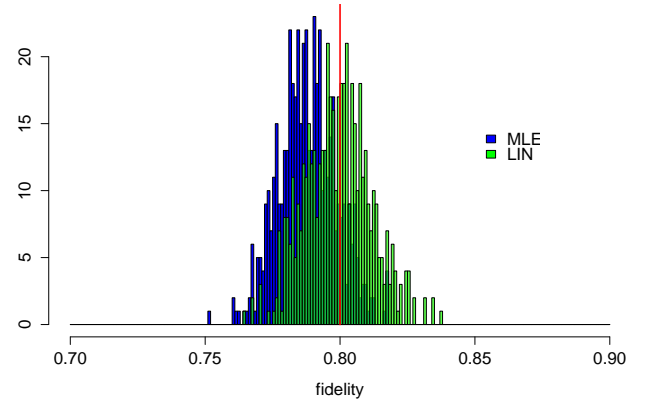


FIG. 2: Fidelity histograms of simulated quantum-state tomography experiments for the noisy GHZ state from Ref. [1]. The true state is a GHZ state with added white noise so that its fidelity is 0.8 with respect to the original target GHZ state.

also take the GHZ state itself as the true state. Then the true fidelity with respect to the target state takes the maximal value of 1. In this situation, we find that all the fidelity values estimated using LIN give the true value of 1, while the MLE values slightly underestimate the fidelity (see Fig. 3(a); $\bar{F}_{\text{MLE}} = 0.999 \pm 0.00036$, the mean and the standard deviation obtained from 500 data sets). That LIN always gives the correct fidelity value, regardless of the statistical fluctuations in the data, is a special feature of the GHZ target state in conjunction with the product Pauli POM [12]. Such a singular situation (as can be seen from more generic examples below) does not provide a fair benchmark from which to draw reliable conclusions about the efficacy of the estimation procedures.

As the second example, we use a different target state, also taken from Ref. [1], namely the four-qubit W-state, $|\text{W}_4\rangle = \frac{1}{2}(|0001\rangle + |0010\rangle + |0100\rangle + |1000\rangle)$. To illustrate the problems of a nonphysical $\hat{\rho}_{\text{LIN}}$, we again consider the extreme scenario where the true state is exactly the target state so that the true fidelity is 1. LIN gives fidelity values ($\bar{F}_{\text{LIN}} = 0.999 \pm 0.015$) that are distributed around the true value but have a wide spread, with a large fraction of the values exceeding 1, a situation without physical meaning. The results obtained via MLE ($\bar{F}_{\text{MLE}} = 0.998 \pm 0.00042$) again slightly underestimate the fidelity, but not by much, and the spread of values is much narrower than for LIN; see Fig. 3(b). That the MLE underestimates the true fidelity value is unavoidable because $F \leq 1$ for physical states and this demonstrates our earlier point that one can only approach an extremal value from one side.

The GHZ and W states considered thus far are quite special states, and particularly so with respect to the product Pauli POM. One wonders about the behaviour for a generic target state. To this end, we randomly pick a pure target state $|\phi\rangle$ and a random true state ρ with a target fidelity close to 1. The selected true state has target fidelity of 0.981; the fidelities calculated via LIN give $\bar{F}_{\text{LIN}} = 0.982 \pm 0.018$ and those gotten via MLE give $\bar{F}_{\text{MLE}} = 0.957 \pm 0.0080$; see Fig. 3(c). Again, a significant fraction of fidelity values obtained from LIN are

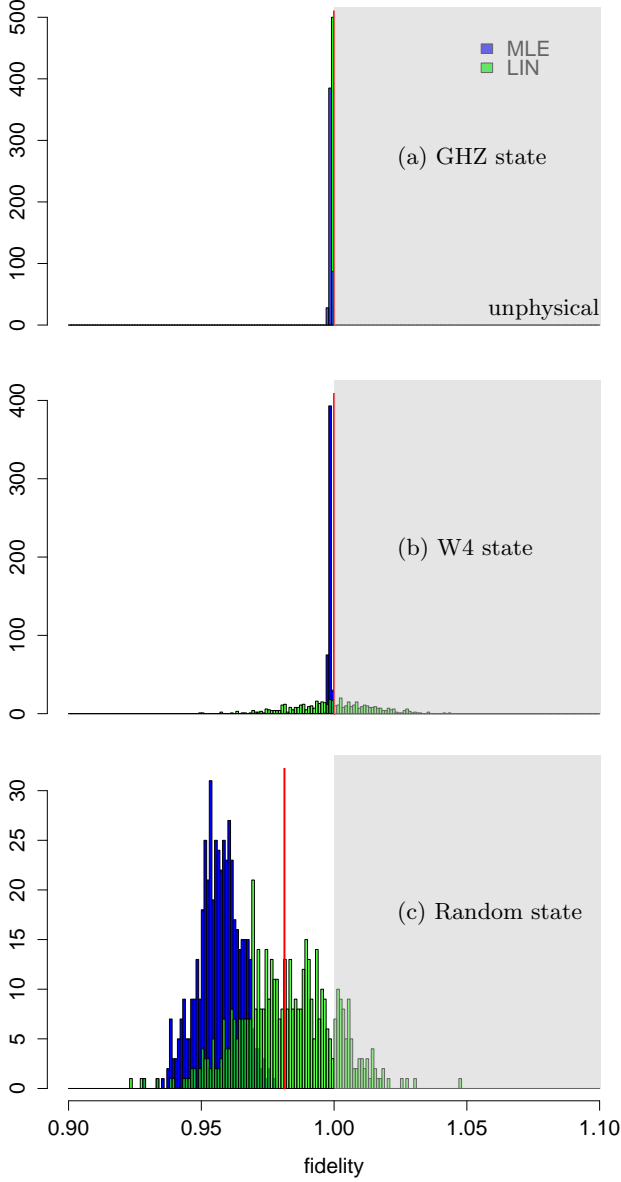


FIG. 3: Fidelity histograms of simulated quantum-state tomography experiments for various states (indicated in plots). The GHZ and W4 states of plots (a) and (b) have true fidelity of 1; plot (c) is for a true state that has fidelity 0.981 with a randomly chosen pure target state. In plots (b) and (c), the histograms for the LIN values extend into the unphysical range of fidelity values exceeding unity.

greater than 1; no such problems appear in the MLE case, by construction, and the MLE range of values has a smaller spread.

The same problems of unphysical fidelity values appear at the other end of the allowed fidelity range, *i.e.*, for F close to 0. We again use the same randomly generated pure state $|\phi\rangle$ as the target state, but now choose a true state with a target fidelity of 0.016. This is the situation in which one's intent is to prepare a state orthogonal to the target state. In this setting, we find that the fidelities calculated via LIN ($\bar{F}_{\text{LIN}} = 0.015 \pm 0.017$) still center around the true value,

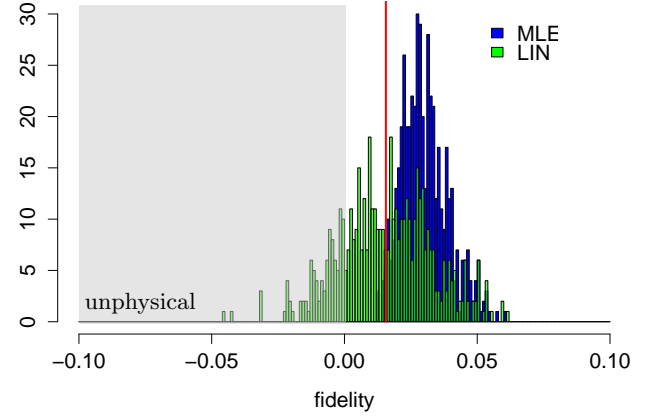


FIG. 4: Fidelity histograms of simulated quantum-state tomography experiments for the same noisy random pure state as in Fig. 3(c), but with the true fidelity being 0.016. The histograms for the LIN values extend into the unphysical range of negative fidelity values.

but with a large fraction being less than 0, another situation that is physically meaningless. On the other hand, the values obtained via MLE ($\bar{F}_{\text{MLE}} = 0.030 \pm 0.0088$) in this case slightly overestimate the fidelity, but again the spread of MLE values is smaller than the spread of LIN values; see Fig. 4.

In all the examples above, LIN suffers from the problem of having unphysical fidelity values. Note that any attempt to fix these larger-than-1 or smaller-than-0 F values from LIN, *e.g.*, set all values > 1 to be equal to 1, will bias the originally unbiased LIN estimator. In all cases, LIN indeed gives a *mean* fidelity closer to the true value (in fact it should be exactly equal, if not for the fluctuations from only 500 runs) than MLE but, apart from the singular case of the GHZ state, has a significantly larger spread than the MLE values, suggesting a possibly larger, or at least comparable, MSE. As mentioned in the previous section, for a single run of the experiment, what matters is not the mean value over many repeated runs, but the MSE; a large variance means that one has a high chance of winding up rather far away from the true value. Thus a meaningful comparison between MLE and LIN requires us to look at the MSE for a variety of states.

Table I presents the results of such a comparison. The first six rows of the table give the variance, bias and MSE values for the examples described above; the remaining rows give the values for a further three randomly chosen target states, with the true state in each case being the target state with added white noise to yield a true target fidelity of 0.8. In all cases, LIN has zero bias (by construction), with the small deviations indicated in the table attributable to finite statistics; however, it has a variance comparable to that from the MLE in every single case. For the randomly chosen states, the MSE values from LIN are slightly smaller than those from MLE, but again, MLE is not designed to reduce the MSE (neither bias nor variance), and sacrificing physicality in LIN is a very high price to pay for the occasional small reduction of the MSE.

TABLE I: The MSE as well as the respective variance and bias in different examples with pure target states. The first column states the ket of the four-qubit target state used in the simulation, with respect to which the true state has the fidelity F_0 of the second column. The kets $|\phi\rangle$, $|\gamma\rangle$, $|\tau\rangle$, and $|\theta\rangle$ are randomly chosen examples. In the third, fourth, and fifth columns, the double entries report the LIN value on the left of the $|$ symbol and the MLE value on the right.

Target	F_0	Mean squared error		Variance		Bias	
		LIN MLE		LIN MLE		LIN MLE	
$ \text{GHZ}_4\rangle$	0.8	1.488×10^{-4}	2.623×10^{-4}	1.480×10^{-4}	1.148×10^{-4}	8.236×10^{-7}	1.475×10^{-4}
$ \text{GHZ}_4\rangle$	1.0	0.0	1.904×10^{-6}	0.0	1.270×10^{-7}	0.0	1.777×10^{-6}
$ \text{W}_4\rangle$	1.0	2.323×10^{-4}	2.642×10^{-6}	2.316×10^{-4}	1.792×10^{-7}	6.663×10^{-7}	2.463×10^{-6}
$ \phi\rangle$	0.981	3.082×10^{-4}	6.681×10^{-4}	3.078×10^{-4}	6.449×10^{-5}	3.812×10^{-7}	6.036×10^{-4}
$ \phi\rangle$	0.016	3.044×10^{-4}	2.937×10^{-4}	3.043×10^{-4}	7.704×10^{-5}	7.918×10^{-8}	2.166×10^{-4}
$ \phi\rangle$	0.8	3.472×10^{-4}	5.649×10^{-4}	3.464×10^{-4}	1.418×10^{-4}	7.934×10^{-7}	4.231×10^{-4}
$ \gamma\rangle$	0.8	2.982×10^{-4}	5.614×10^{-4}	2.982×10^{-4}	1.568×10^{-4}	9.143×10^{-10}	4.046×10^{-4}
$ \tau\rangle$	0.8	3.410×10^{-4}	5.087×10^{-4}	3.410×10^{-4}	1.320×10^{-4}	3.371×10^{-9}	3.767×10^{-4}
$ \theta\rangle$	0.8	3.758×10^{-4}	5.859×10^{-4}	3.758×10^{-4}	1.562×10^{-4}	4.369×10^{-8}	4.297×10^{-4}

V. CONCLUSION

The method of linear inversion may appear simple and intuitive—after all, relative frequencies are asymptotically the correct frequentist’s interpretation of probabilities. However, many difficulties arise, both in the construction of $\hat{\rho}_{\text{LIN}}$ (having to put in ad-hoc fixes for internal consistency), as well as in the interpretation of the estimated quantities (if one can compute them at all from $\hat{\rho}_{\text{LIN}}$, and having to deal with parameter values of no physical meaning even when computation is possible). Both these problems disappear when insisting on physical estimators. This requirement is natural in all cases, since the true state must be physical, not just asymptotically so. Even if one is willing to deal with the problems of having a nonphysical estimator, reducing the MSE, not just the bias, is the key, and the reduction had better be substantial enough for one to want to cope with the nonphysicality issues.

To end the discussion, we remind the reader of a point briefly mentioned in the opening paragraphs: If one is interested in estimating only a single parameter, a direct estimation from the data without first going through an estimator for the state generally works better. Thus, our assessment here of the

LIN versus MLE is also lacking in that we should compare their performance for a variety of parameters computed from $\hat{\rho}_{\text{LIN}}$ and $\hat{\rho}_{\text{MLE}}$, for one would not report these estimators unless the interest is in a few different parameters computed from the state. Yet, given that LIN already does poorly in terms of target fidelity, it is hardly necessary to provide more evidence against the use of LIN by exploring other parameters. Furthermore, a more comprehensive summary of the data is provided not by the *point* estimators $\hat{\rho}_{\text{LIN}}$ and $\hat{\rho}_{\text{MLE}}$, but by regions of estimators, which can circumvent some of the problems highlighted here; see [13] for confidence regions, and [14] for credible regions.

Acknowledgments

We thank Yong Siah Teo and David Nott for stimulating discussions. The Centre for Quantum Technologies is a Research Centre of Excellence funded by the Ministry of Education and the National Research Foundation of Singapore.

-
- [1] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne, eprint arXiv:1310.8465 [quant-ph] (2013).
 - [2] Z. Hradil, Phys. Rev. A **55**, R1561 (1997).
 - [3] M. G. A. Paris and J. Řeháček, eds., *Quantum State Estimation*, Lect. Notes Phys. **649** (Springer-Verlag, Heidelberg, 2004).
 - [4] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).
 - [5] This F is actually the square of the fidelity, if a close analog of

the fidelity of two probability distributions is wanted. We here adopt the convention of Ref. [1] and use the term “fidelity” for F thus defined.

- [6] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Phys. Rev. A **70**, 052321 (2004).
- [7] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, J. Math. Phys. **45**, 2171 (2004).
- [8] Note that splitting up a single set of data into very many smaller sets and hence very many estimates does not help as the smaller

sample sizes will result in a larger variance.

- [9] E. L. Lehmann and G. Casella, *Theory of Point Estimation* (Springer Texts in Statistics) 2nd ed. (Springer-Verlag, New York, 2003).
- [10] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).
- [11] What is particularly ironic is Jaynes's point that, for his example, the estimator that minimizes the MSE actually requires one to "double the bias" by dividing the variance by $N + 1$ (for sample size N) instead of the usual unbiased correction of $N - 1$, as in Ref. [37] of Ref. [1].
- [12] If the GHZ state is the target, then the target fidelity involves the expectation values of $[XXXX]$, $[YYYY]$, $[ZZZZ]$ as well as $[XXYY]$ and $[ZZ\omega\omega]$ and the observables obtained by permuting the four qubits. The GHZ state is an eigenstate of every one of these 15 observables. Hence, there are no statistical fluctuations when their expectation values are estimated from the data obtained from the product Pauli POM.
- [13] M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012); R. Blume-Kohout, arXiv:1202.5270 [quant-ph] (2012).
- [14] J. Shang, H. K. Ng, A. Sehwat, X. Li, and B.-G. Englert, New J. Phys. **15**, 123026 (2013).